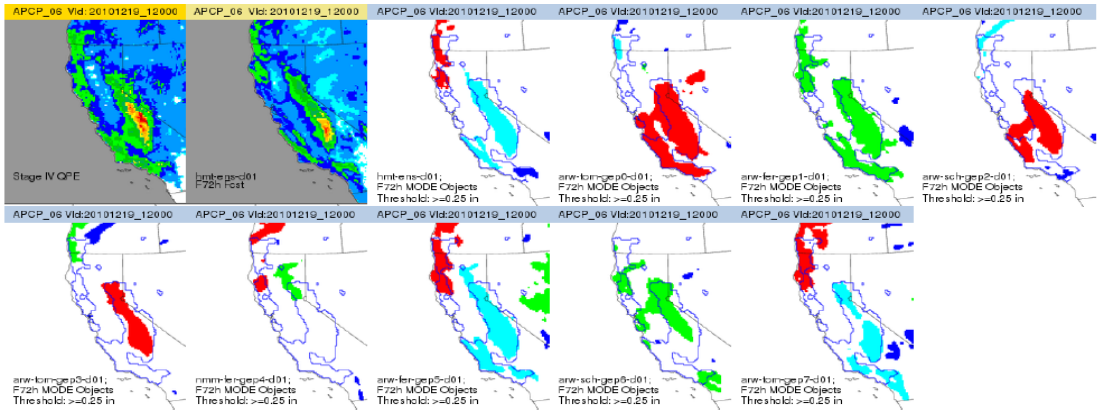


Verification and Diagnoses of Ensemble QPF Forecasts during Extreme Events in California during the HMT Winter Exercises

*Edward Tollerud[1,2], **Tara Jensen[1,3]**, John Halley Gotway[1,3], Paul Oldenburg[1,3], Wally Clark[2] , Tressa Fowler[1,3] , Stanislav Stoytchev [1,2], Barb Brown[1,3], Ellen Sukovich[2], Randy Bullock[1,3] , and Isidora Jankov [4]*

[1]Developmental Testbed Center (DTC)
[2]Earth Systems Research Laboratory, NOAA, [3] Research Applications Laboratory, NCAR,
[4] CIRA/Colorado State University

3rd NOAA Testbeds and Proving Ground Workshop, 1-3 May 2012, Boulder, CO

Developmental Testbed Center

DTC/HMT Collaboration Goals

✓ **Evaluation and Diagnoses for HMT-West Ensemble Forecasts** of Extreme Precipitation Events (e.g., real-time web product for HMT)

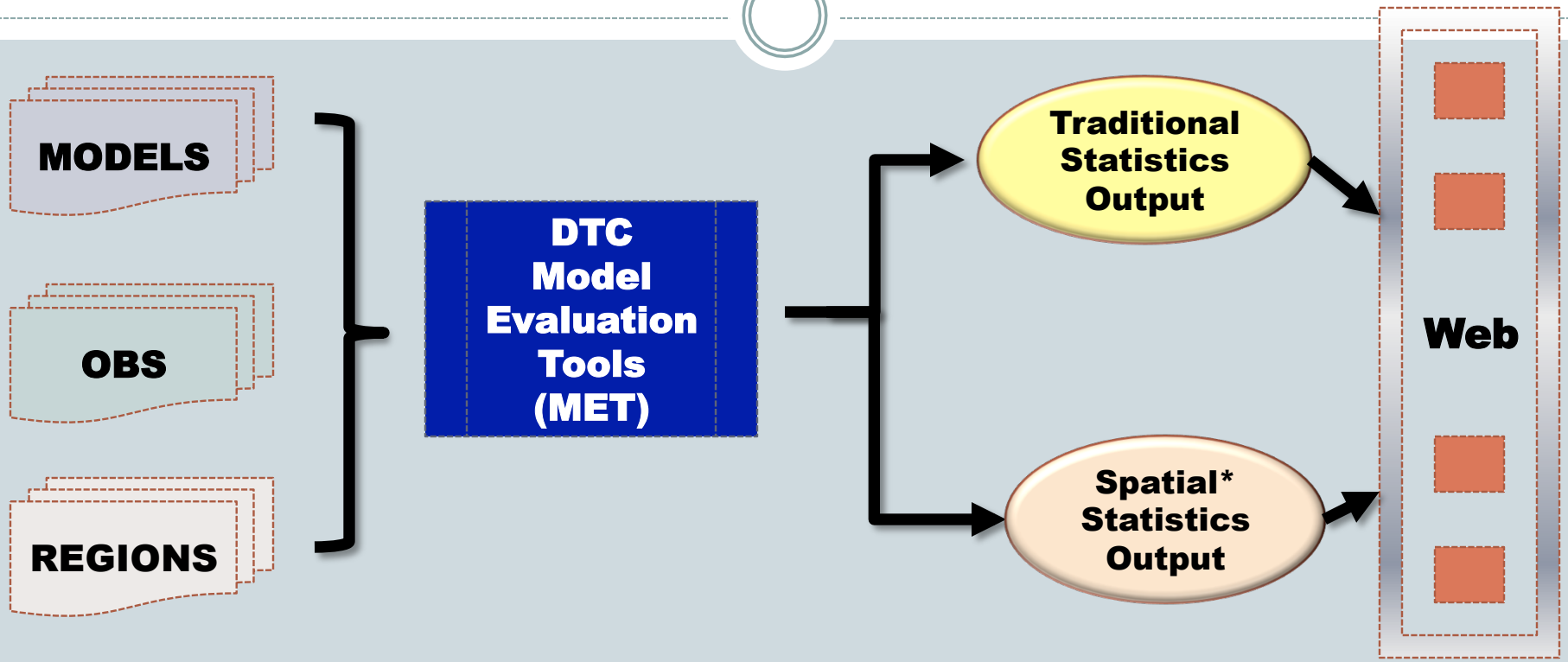
✓ **Motivate, Develop, and Evaluate new verification strategies** (MET, MODE, and METViewer in particular; e.g., roc, auc, rank histogram, performance diagram,…)

✓**Assess Model and Verification Configuration Options** (Resolution, Initialization, Domain, Event Selection, etc.)

✓**Inter-compare Forecasting Systems in high-precipitation scenarios**, including storm-scale research and EMC operational models

✓**Assess Impacts of Verification dataset selection** (analyses, point obs, etc.) – not covered here

Testbed Collaboration Methodology

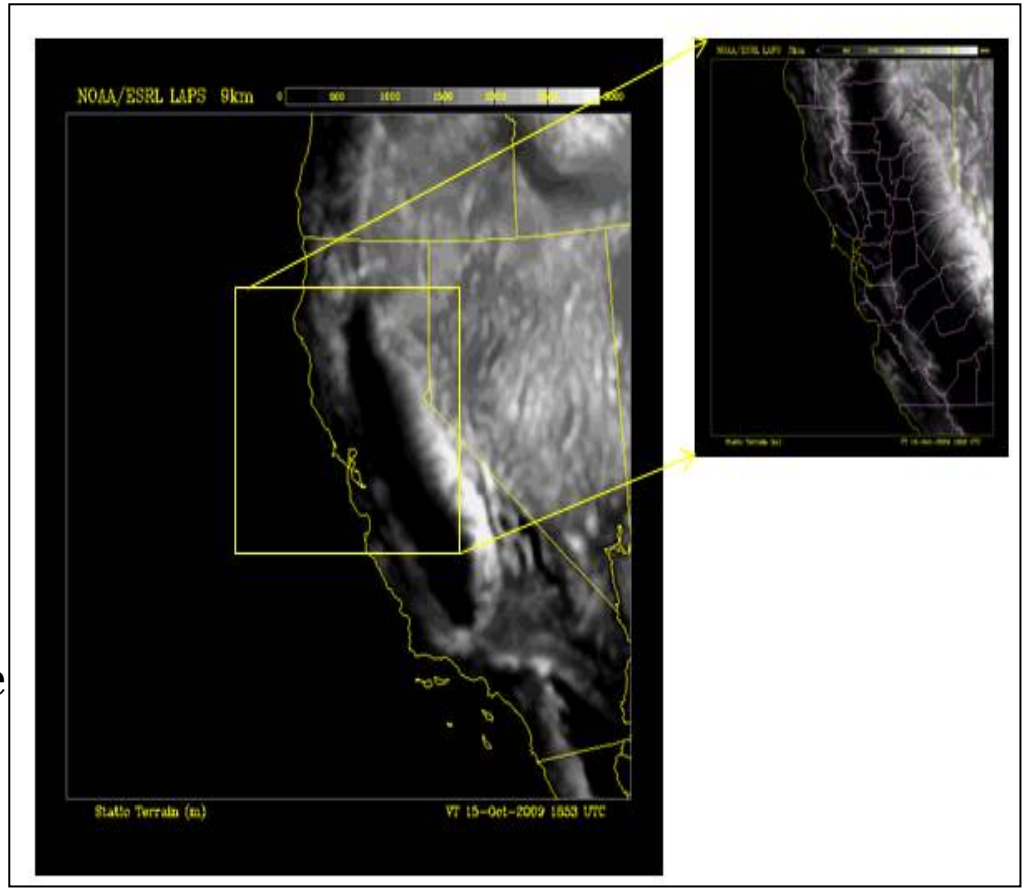


*MODE, Neighborhood, etc

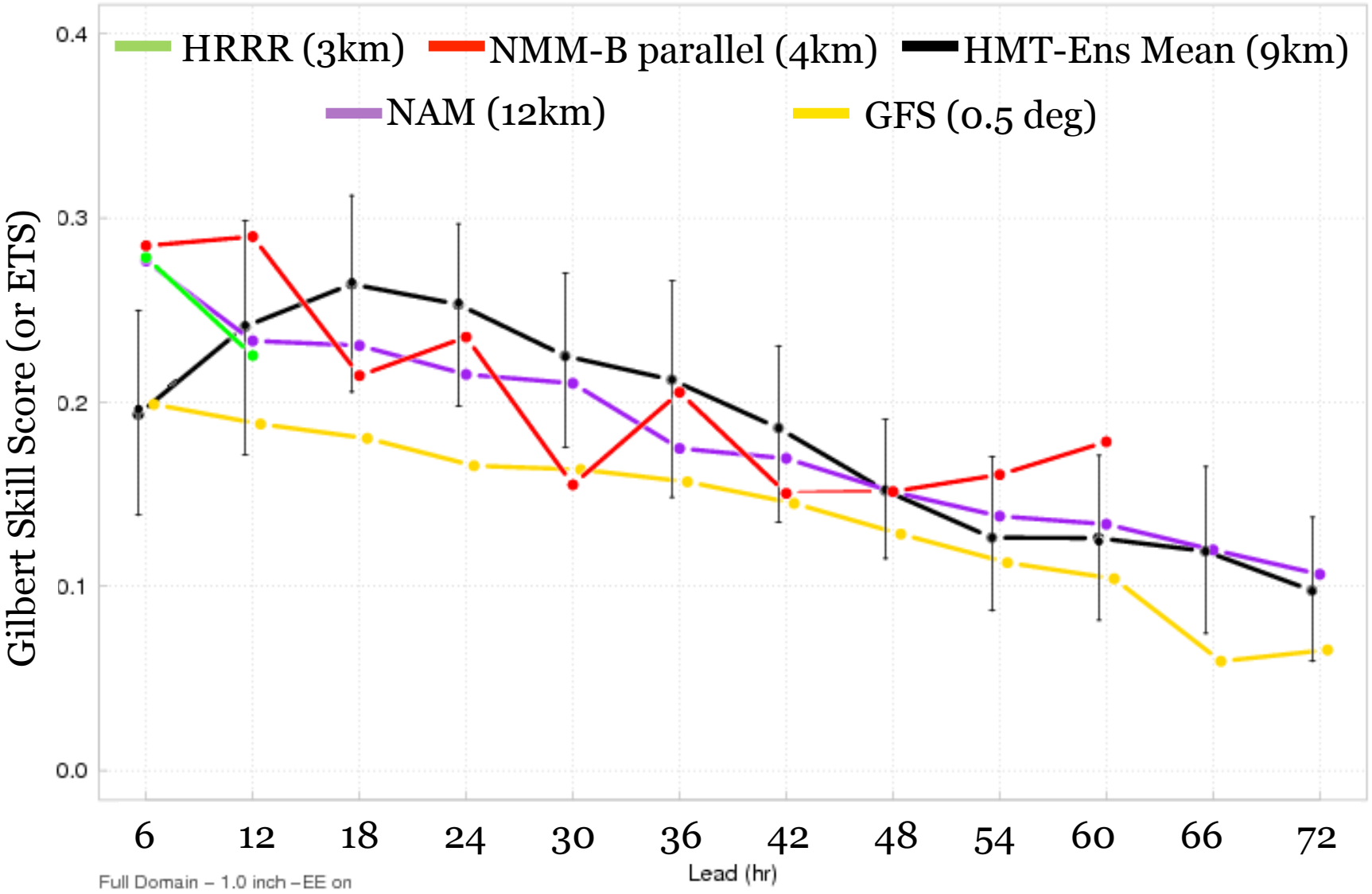
MET is a set of NWP evaluation tools developed by the Developmental Testbed Center (DTC) to help them assess and evaluate the skill of their model predictions. It is ***free to download*** and there is a helpdesk available.

ESRL/GSD and HMT Ensemble Modeling System

- WRF model 8-member ensemble; 1 control
- Outer domain 9km; Nested domain 3 km
- Hybrid members: Multi physics packages, two model cores, and different GFS initial conditions
- Model runs to 5 day lead time; DTC evaluated first 72 hours
- DTC built demonstration real-time web display
- Evaluation focus on QPF with addition of state variables in 2011



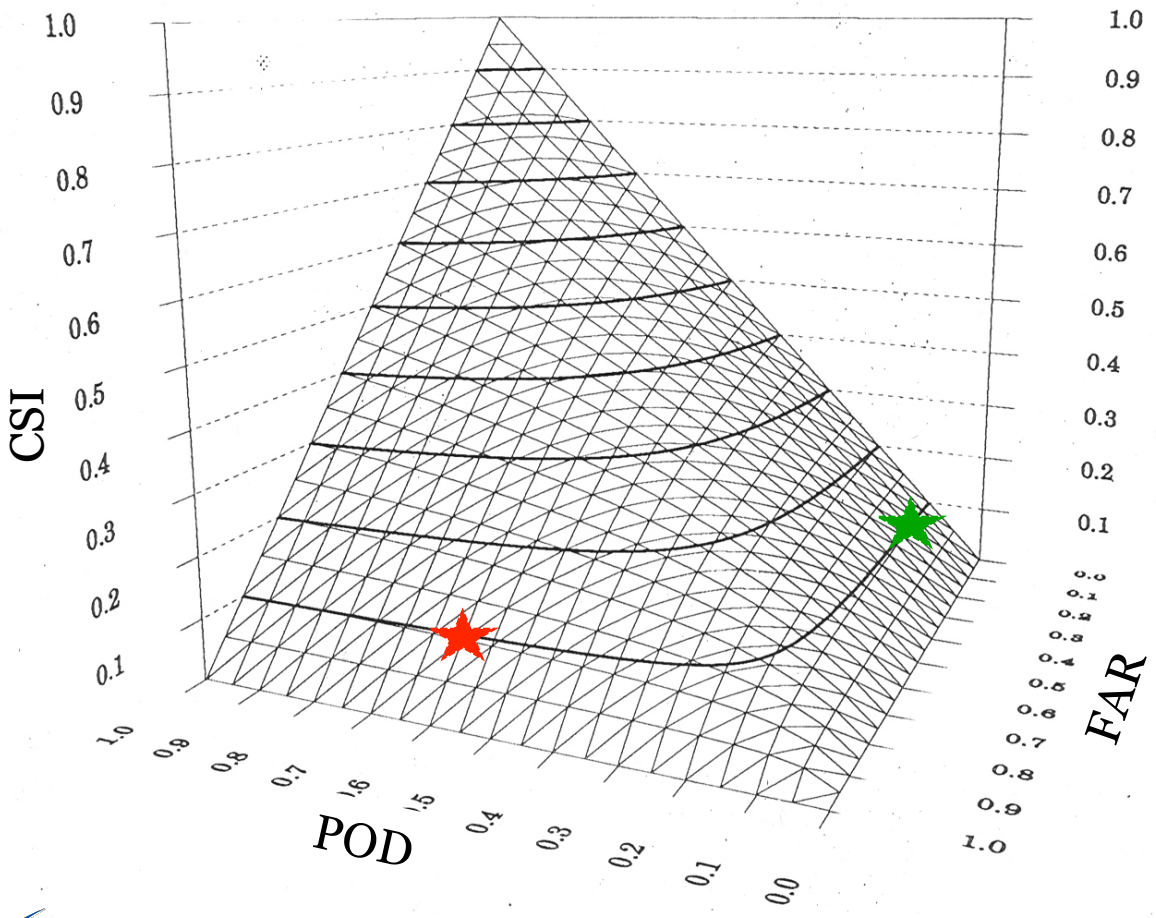
Model Intercomparison for 2010-2011 HMT-West



6hr accum. > 1.0”

Relationships among scores

- CSI is a *nonlinear* function of POD and FAR
- CSI depends on base rate (event frequency) and Bias



$$\mathrm{CSI} = \frac{1}{\frac{1}{\mathrm{POD}} + \frac{1}{1-\mathrm{FAR}} - 1}$$

$$\mathrm{Bias} = \frac{\mathrm{POD}}{1-\mathrm{FAR}}$$

Very different combinations of FAR and POD lead to the same CSI value

HMT Performance Diagram

All on same plot

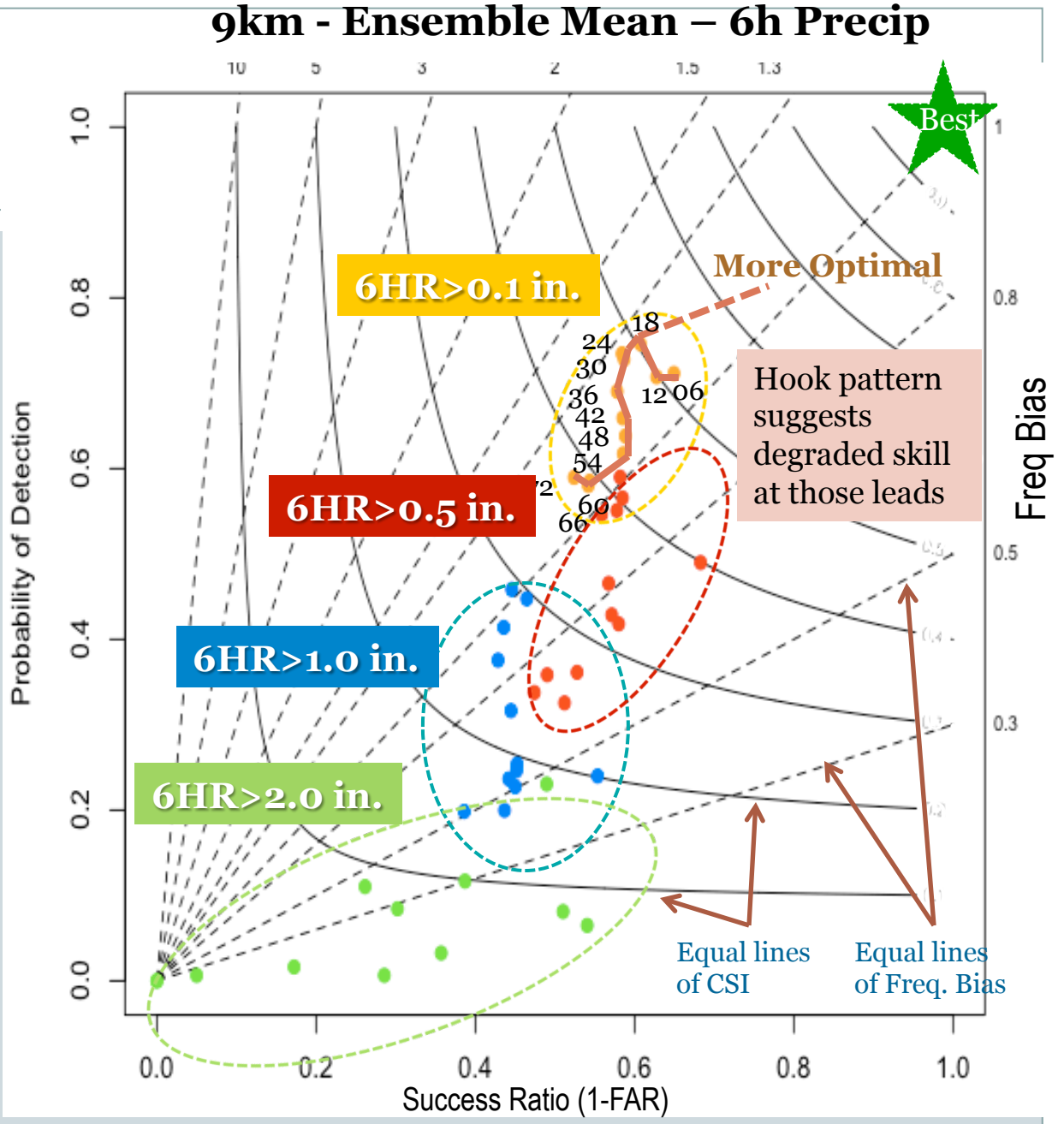
- POD
- 1-FAR (aka Success Ratio)
- CSI
- Freq Bias

Dots: Scores Aggregated Over Lead Time

Colors: Different Thresholds

Here we see:

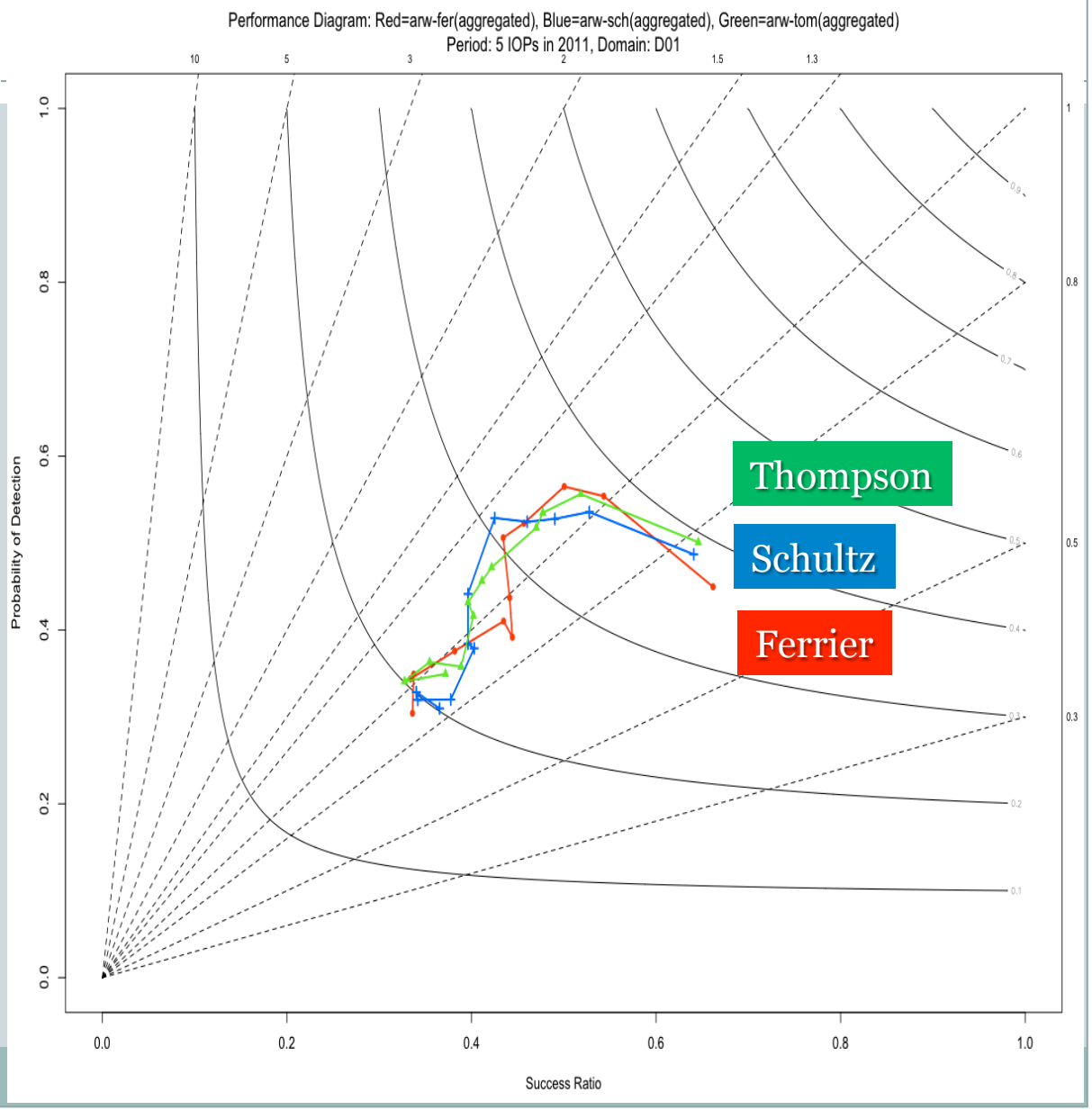
- Decreasing skill with higher thresholds even with multiple metrics
- Highest skill at 18-24h leads



Roberts et al. (2011), Roebber (WAF, 2009), Wilson (presentation, 2008)

Impact of Microphysics on 2010-2011 Results

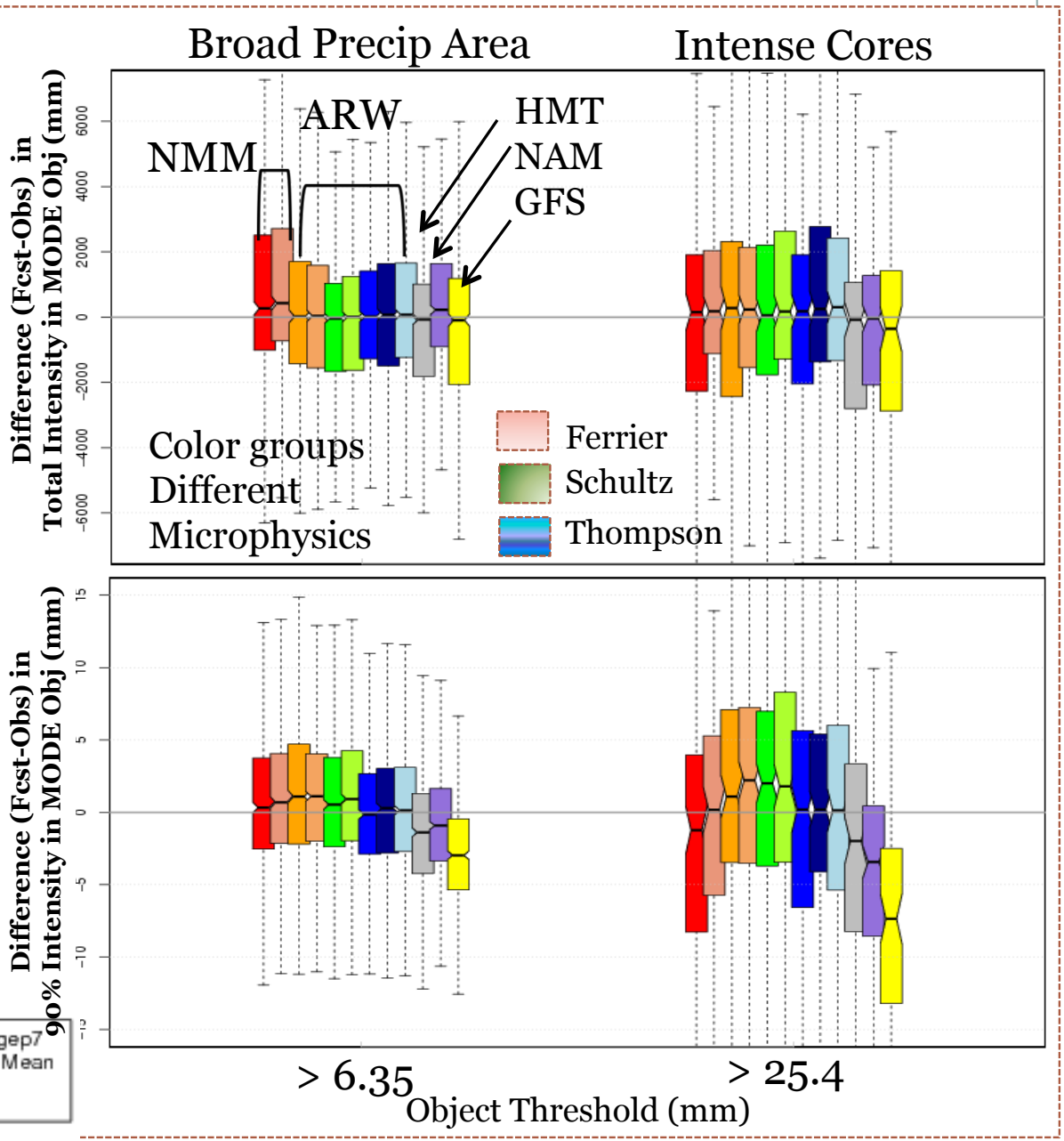
- No systematic microphysics impacts last season
- Performance diagrams similar
- Total Intensity distributions similar for most HMT members
- 90% Intensity show some differences, especially at higher thresholds
- HMT Ens Mean does not have same performance as ind. members

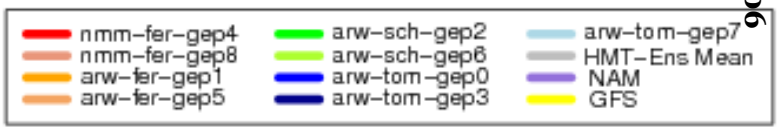


Impact of Microphysics on 2010-2011

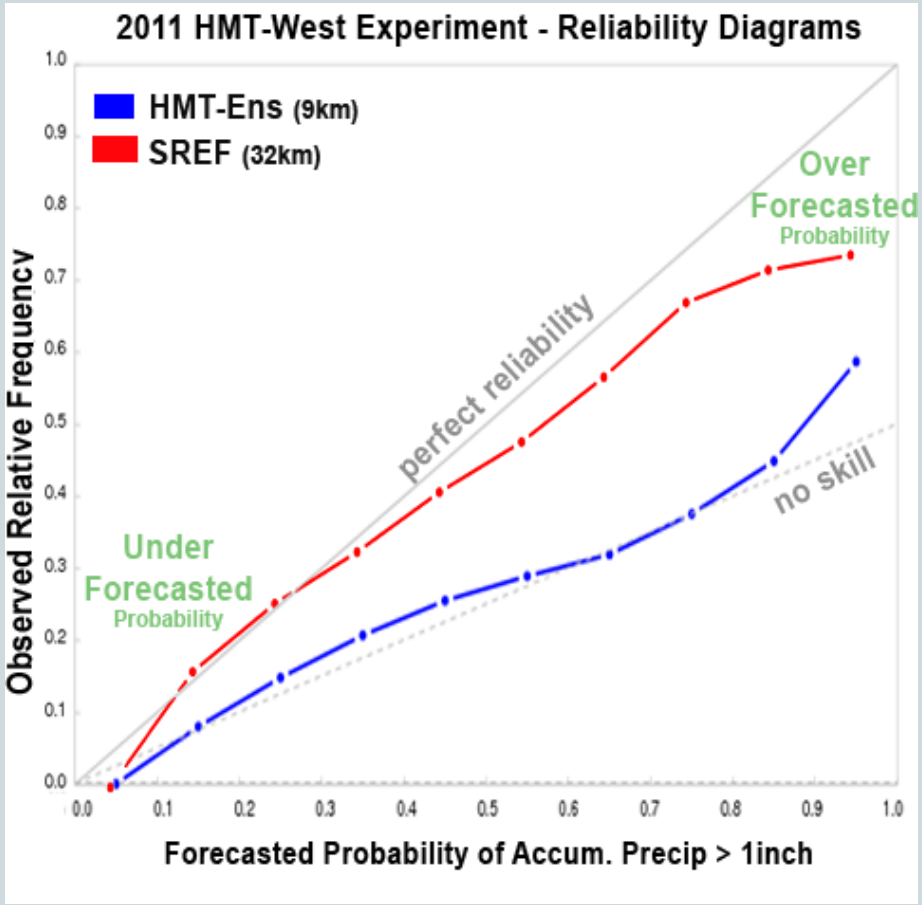
Using Attributes from MODE Objects

- No systematic microphysics impacts last season
- Performance diagrams similar
- Total Intensity distributions similar for most HMT members
- 90% Intensity show some differences, especially at higher thresholds
- HMT Ens Mean does not have same performance as ind. members

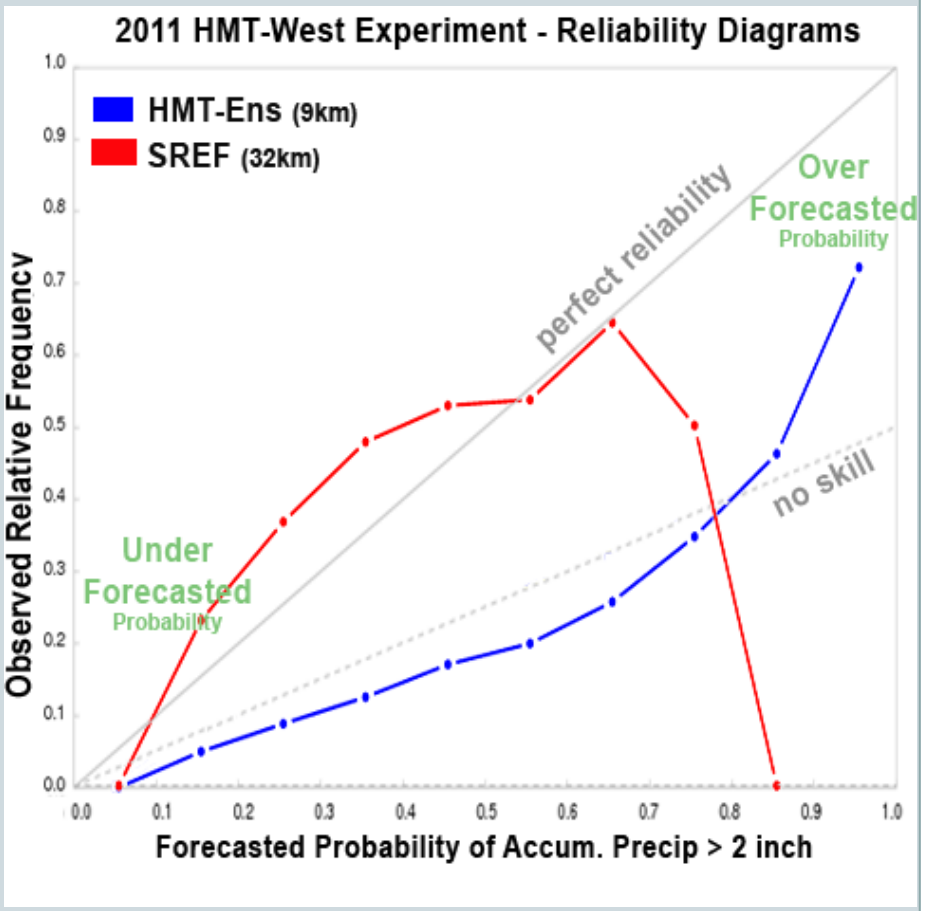




Ensemble Reliabilty



PROB(APCP_06)>1 inch



PROB(APCP_06)>2 inch

Valuable Insights and Lessons Learned: Some gained, Some still in process

- ✓ Resolution improves performance
- ✓ Scores for ensemble means are generally different from the mean score of the ensemble members – understanding how to "ensemble" scores is an area of research
- ✓ Model Core - Microphysical Impacts -Initialization impacts all need more investigation but we are now have a more effective set of tools to do this
- ✓ Performance diagrams may be helpful in diagnosing model performance problems

Year 3 (2011-2012) Season Emphasis

- ✓ Continued evaluation of QPF
- ✓ Expansion to state variable (T, SPFH, U/V, HGT) and critical moisture variables for HMT (IWV, Freezing Level)
- ✓ Inclusion of AFWA Ensemble (at the request of EMC) – Thanks to Evan Kuchera and Scott Rentsler

- ✓ Just finished final evaluation runs of season (yesterday)

- ✓ Will be presenting results at:
  WAF/NWP – CMOS conference at end of May
  WRF Users Workshop – end of June

Thanks for your attention

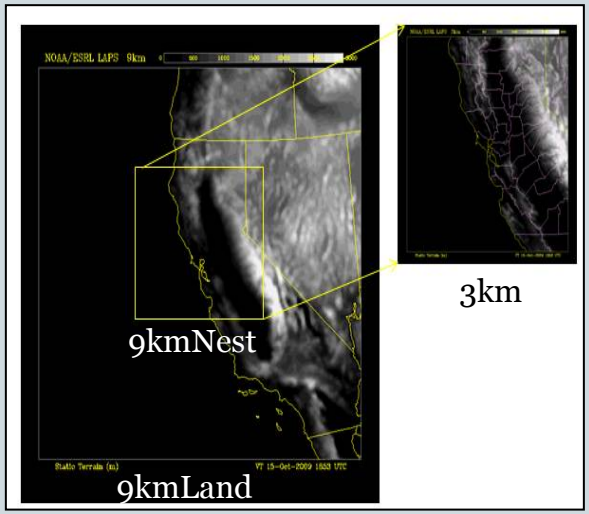
- Thanks to the DTC collaborators: ESRL/GSD, ESRL/PSD, EMC, and AFWA
- This DTC/HMT work was funded by USWRP

<u>For more information</u>

- Edward Tollerud (edward.tollerud@noaa.gov)
- Tara Jensen (jensen@ucar.edu)
- Brian Etherton (brian.etherton@noaa.gov)

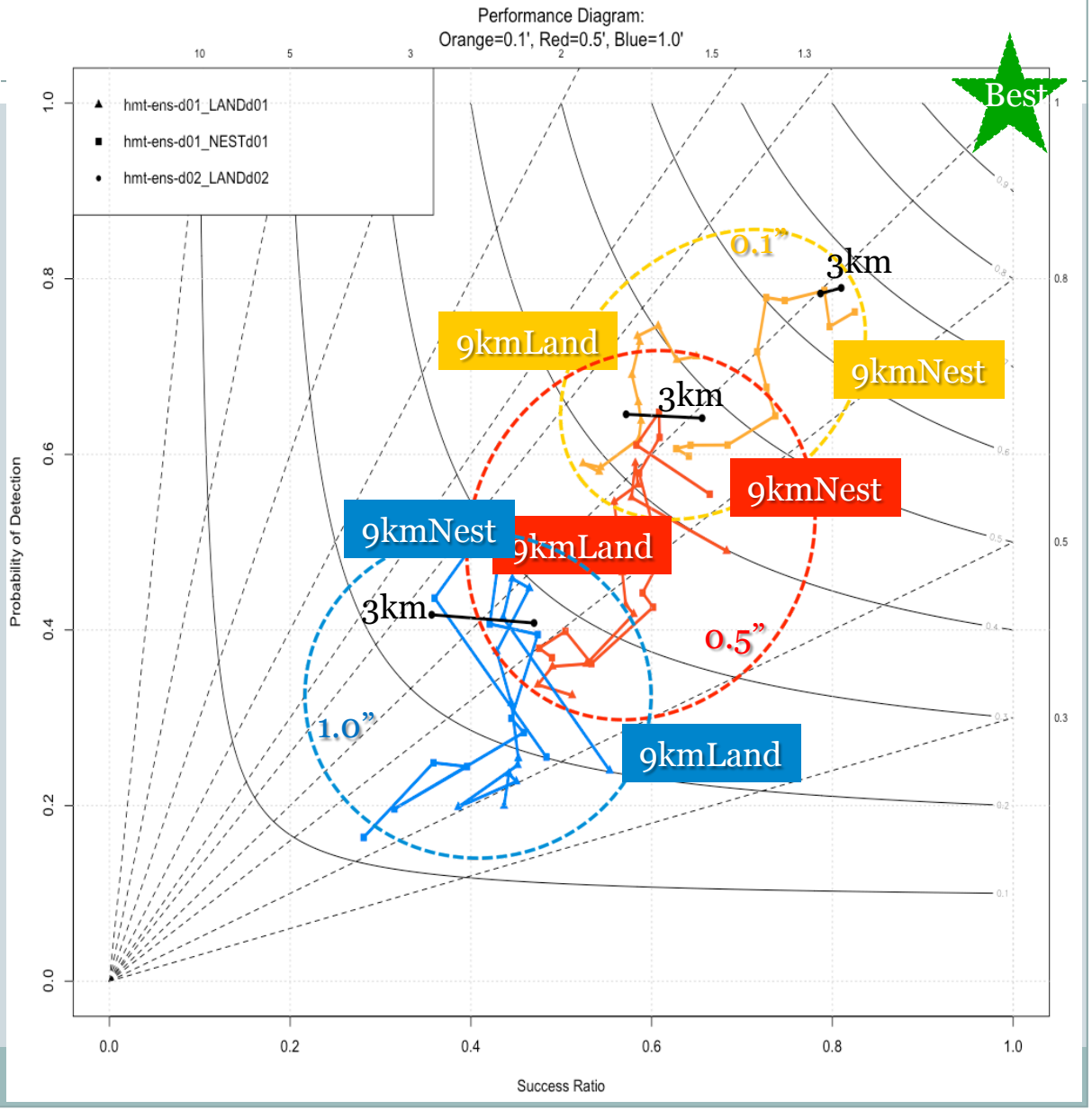
- http:/www.dtcenter.org/eval/hmt

Impact of Domain on 2010-2011 Results

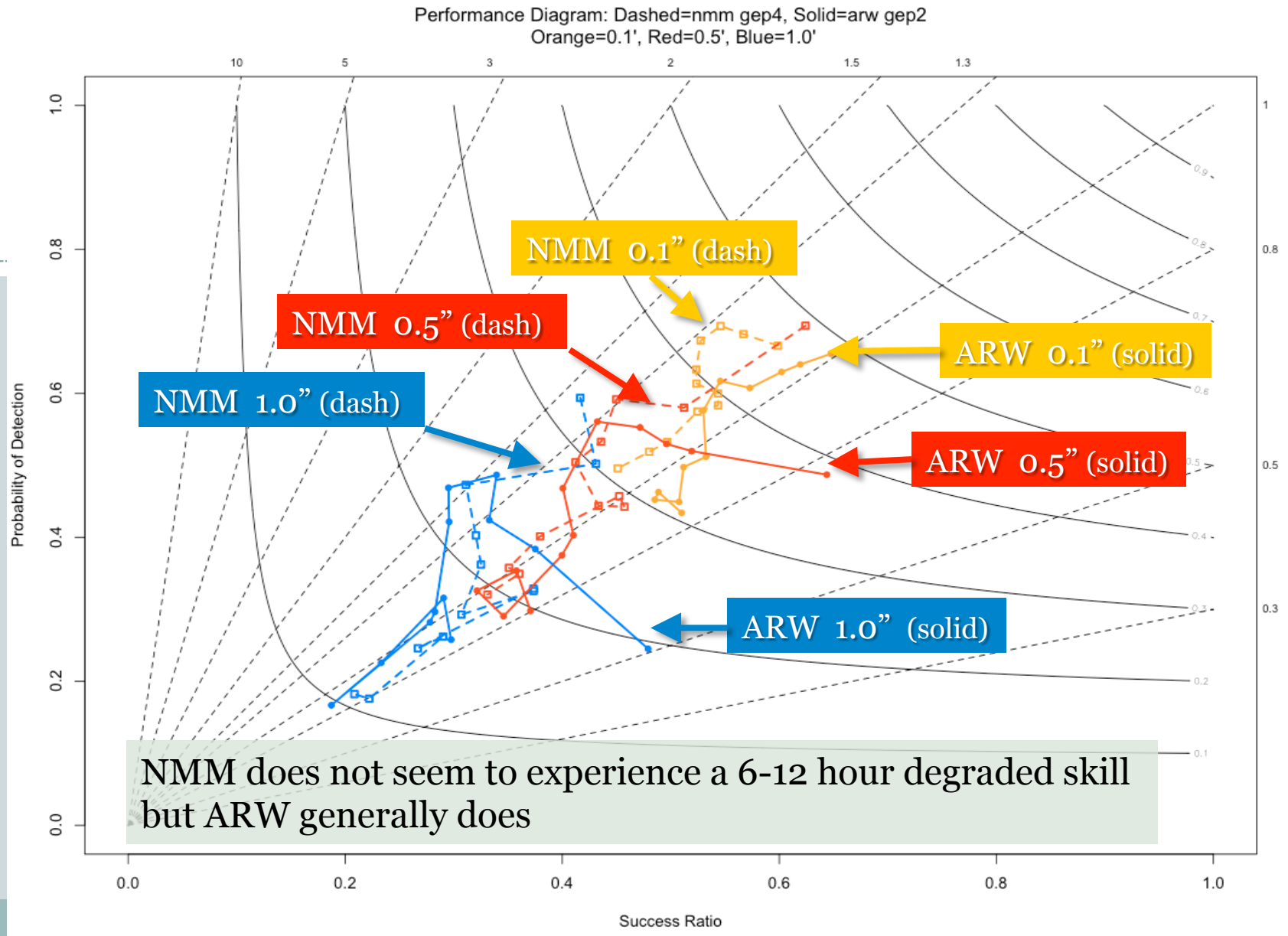


Eval of 9km domain over Nest footprint (9kmNest) appears to have greater skill at short leads

3km domain has more skillful Performance Diagrams at 6-12 hr leads



Impact of Model Cores on 2010-2011 Results



NMM does not seem to experience a 6-12 hour degraded skill but ARW generally does

Area under ROC Curve